



**University of
Zurich**^{UZH}

**Zurich Open Repository and
Archive**

University of Zurich
University Library
Strickhofstrasse 39
CH-8057 Zurich
www.zora.uzh.ch

Year: 2014

The Role of Grammaticality Judgments Within an Integral Approach to Brazilian Portuguese Bare Nominals

Wall, Albert

DOI: https://doi.org/10.1007/978-3-319-05675-3_6

Posted at the Zurich Open Repository and Archive, University of Zurich

ZORA URL: <https://doi.org/10.5167/uzh-106653>

Book Section

Accepted Version

Originally published at:

Wall, Albert (2014). The Role of Grammaticality Judgments Within an Integral Approach to Brazilian Portuguese Bare Nominals. In: Hemforth, Barbara; Mertins, Barbara; Fabricius-Hansen, Cathrine. Psycholinguistic Approaches to Meaning and Understanding across Languages. Cham: Springer, 143-173.

DOI: https://doi.org/10.1007/978-3-319-05675-3_6

[Manuscript accepted for publication in: Hemforth, B.; Mertins, B. & Fabricius-Hansen, C. (eds.) 2014. *Psycholinguistic approaches to Meaning and Understanding across languages*. Cham: Springer. DOI: 10.1007/978-3-319-05675-3_6]

5 The role of grammaticality judgments within an integral approach to Brazilian Portuguese Bare Nominals

Albert Wall

University of Tübingen, SFB833¹

Abstract

This paper is mainly concerned with two points: The first one is a better theoretical foundation of the interpretation of acceptability judgment tasks (AJTs) in studies on Brazilian Portuguese (BrP) bare nominals (BNs). I draw on Bader & Häussler's (2010) model which is based on Signal Detection Theory and show that an (explicitly or implicitly) binary approach to AJTs on BrP BNs fails to capture the whole picture. This is exemplified by contrasting the two AJT studies on specific and definite BNs presented in this paper with other experimental approaches to BrP BNs. The second concern is the status of these rather marginal forms in BrP. It will be claimed that only an approach combining different empirical methods can give a sufficiently clear picture. In order to support this claim, a third experiment, namely an elicitation task will be presented and discussed.

5.1 Introduction

At least since their appearance as counter-evidence to Chierchia's (1998) well-known Nominal Mapping Parameter in Schmitt & Munn (1999)², Brazilian Portu-

¹ This work was financed by the Deutsche Forschungsgemeinschaft, SFB 833, project C3. I would like to thank Sam Featherston, Oliver Bott, Fabian Schlotterbeck and Janina Rado for their support regarding technical implementation, statistical analysis and much constructive feedback. Two anonymous reviewers helped to improve the paper at several critical points. I am also grateful to all the people in Brazil who made the data collection possible, special thanks to José Simões da Silva, Célia Regina dos Santos Lopes and Bruno Festas. The interpretation of the results as well as remaining errors are completely my responsibility.

² For a recent discussion of Schmitt & Munn's claim with regard to "semantic parameterization" and the role of the BrP data as counter-evidence to Chierchia (1998), see Dayal (2011: 1103ff) and references therein.

guese (BrP) BNs have received a considerable degree of attention from researchers interested in the syntax and semantics of the nominal domain. This is due to the fact that BrP, which has a full-fledged Romance-style article paradigm, at the same time allows for plural as well as singular BNs in argument position in surprisingly many contexts. In the words of Braga et al. (2010: 75f): “[...] the determiner phrase [...] ranges from overtly definite and indefinite phrases, both singular and plural, to a complete system of bare noun phrases in argument position: bare singulars [...], bare plurals [...] and bare mass [...]”. The investigation of this topic has led to a very dynamic and productive but also controversial discussion during the last decade, especially in the case of the so-called “bare singulars” (BSs)³, to which this paper is mainly devoted, as well. Examples (1) and (2) contrast a BS with a bare mass noun:

- (1) Professor trabalha muito.
Teacher works much
‘Teachers work a lot.’ (Müller & Oliveira 2004: 12)

- (2) Leite faz bem pra saúde.
milk makes well for health
‘Milk is healthy.’ (Braga et al. 2010: 76)

While (2) is not surprising at all from a Romance (or Germanic) perspective, structures like (1) would not be expected. Nonetheless, (1) can be taken as the natural way to express genericity in Brazilian Portuguese. Languages like English would require a (bare) plural in such cases. The BrP BS also occurs in other contexts and may have different interpretations from the one exemplified above. The relevant examples will be introduced in the course of the discussion. As will be shown in the subsequent sections, the proponents of different theoretical approaches not only diverge in the analysis of these forms, but also with respect to which sentences containing BS arguments they consider as grammatical.

This is not the place to give a detailed overview of the different theoretical approaches and the respective claims, since the main focus will not be on the theoretical implications of the particular views in the first instance but rather on the prominent role diverging grammaticality judgments play in this debate and the problems that arise thereby. For that reason, the claims from the different approaches will be introduced only as far as they are necessary for the subsequent discussion. In this section, I will first briefly describe the relevant data from BrP and review the main controversies in the recent literature on BrP BNs which have triggered experimental studies, as well as the experimental work done so far with the intention to clarify those issues: the question whether BSs are mass nouns and whether they may denote kinds (5.1.1 & 5.1.2). It will become clear that while in

³ This term is nothing more than a descriptive label, indicating the absence of plural morphology (*bare nominals* might also include bare plurals).

both cases the authors use Acceptability Judgment Tasks (AJTs) in order to separate the grammatical sequences from the ungrammatical ones, the obtained data pose a problem for interpretation since they show a considerable amount of variation and are not very clear-cut. In my view, this problem is not properly addressed in the previous experimental studies. In order to overcome this deficit, I will introduce Bader & Häussler's (2010) model of the relationship between gradient acceptability scores and a binary concept of grammaticality as a basis for the overall discussion (5.1.3). Then, I turn to a phenomenon either neglected or simply declared ungrammatical or non-existent in most of the literature, namely definite/specific interpretations and anaphoric uses of BSs (5.1.4).

In section 5.2 I present three experiments which address questions concerning the definite/specific uses of BSs (e.g. their acceptability and conditions of use). They also illustrate some of the problems that must be dealt with in order to interpret gradient grammaticality judgments. Finally, they give first clues about the grammatical status of the specific/definite examples and their distribution.

The theoretical implications of the results are discussed in section 5.3: Firstly, the grammatical status of the specific/definite BSs will be addressed (5.3.1). I will claim that at the moment, Pires de Oliveira & Rothstein's (2011, in press) approach is the one which, besides having other advantages, can most straightforwardly accommodate the results of the experiments which confirm the existence of the phenomenon. Secondly, in 5.3.2, I will discuss the role of AJTs within an integral approach to BrP BNs and argue that (i) marginal phenomena should not be neglected in linguistic theorizing and (ii) that only a combination of different methods can provide a clear picture. Section 5.4 summarizes the conclusions and gives an outlook on questions and tasks for future work.

5.1.1 Empirical controversy 1: Are BSs mass nouns?

According to Schmitt & Munn (1999) and further co-authored publications, BSs (which are taken to be count nouns since they can be pluralized) and bare mass nouns are syntactically and semantically distinct because only the former can be subjects of individuating predicates (predicates which "select for atoms" (cf. fn. 7), reflexives, reciprocals), while the latter can't. Consider (3) vs. (4):

- (3) Criança pesa 20 quilos nesta idade.
child weighs 20 kilos in+this age
'Children weigh 20 kilos at this age.' (Schmitt & Munn 1999: 348)
- (4) *Ouro pesa duas gramas.
gold weighs two grams
'*Gold weighs two grams.' (Schmitt & Munn 1999: 348)

Similar facts can be demonstrated with reflexives and reciprocals, which also reject mass subjects in languages like English. These facts lead Schmitt & Munn to postulate differences in the syntax of BrP BSs and mass nouns⁴, namely in the make-up of the so-called “functional projections”, i.e. the postulated structures in the syntactic tree of a nominal phrase (NP) “above” the noun. However, there are also reasons to treat BSs and bare mass nouns alike, as e.g. Braga et al. (2010) correctly point out. They argue that the apparent contrast explored by Schmitt & Munn arises because those authors only consider special cases of count and mass nouns, namely “atomic count nouns”, such as *child* versus “non-atomic mass nouns” like *gold*, which are the prototypical cases but by no means the only ones: There are “atomic” mass nouns such as *furniture* and “non-atomic” count nouns such as *fence*. In these cases, the individuating predicates tend to reject the count noun, i.e. the BS:

- (5) Móbiília (dessa marca) combina uma na outra.
 Furniture (of+this brand) fits one in+the other.
 ‘Furniture (of this brand) fits each other.’ (Braga et al. 2010: 79)
- (6) ?? Cerca (nesse terreno) mede 3 metros.
 Fence (in+this land) measures 3 meters. (Braga et al. 2010: 79)

Braga et al. report two AJTs: one with the aim to show that here the predictions of Schmitt and Munn’s approach break down and a second one in order to test whether it is actually the “natural atomicity”⁵ of the noun that matters for individuating predicates. Since the second experiment largely reproduces and confirms the results of the first experiment, I will only discuss the second one here.

In the experiment, sentences parallel to (3) - (6), featuring different individuating predicates were tested for acceptability. The 12 target sentences contained the following types of subject nouns which were treated as conditions in the analysis:

- naturally atomic (NA)/count
- naturally atomic (NA)/mass
- non-naturally atomic (NNA)/count
- non-naturally atomic (NNA)/mass

The material also was controlled for length of sentences (equal number of words, 5-8 syllables per sentence), and equal number of items per condition. The target sentences were divided on four lists and mixed with 6 distracters. Each list of this pencil & paper task was judged by 50 participants (university students, none from linguistics). For the judgment, a 15cm long line was provided next to

⁴ For a detailed review of Schmitt & Munn’s original claims, see Braga et al. (2010).

⁵ Braga et al. take this term from Rothstein (2010) where it refers to inherent individuality in contrast to “formal atomicity” (being an atom in a Boolean structure) and “semantic atomicity” (atomicity relative to a certain context).

each of the items, with a smiling face and at its left end and a sad one at its right end. The question to be answered was: “does the sentence sound fine?” (Braga et al. 2010: 84), and the participants were asked to mark the answer on the line between “yes” (smiling face) and “no” (sad face).

For the analysis of the results, the line between the two faces was divided into three equally large regions: one close to the smiling face, one close to the sad face and the remaining area in the middle. The regions were interpreted as “acceptable”, “unacceptable” and “undecided”, respectively. For each region and condition, the number of ratings was summed up and the distribution of the judgments was compared. Table 1 provides the results (12 target sentences with 50 judgments each result in 600 judgments, percentages are given by condition):

condition	“acceptable”	“undecided”	“unacceptable”	total
NA-count	67 (≈45%)	16 (≈10%)	67 (≈45%)	150
NA-mass	47 (≈31%)	26 (≈17%)	77 (≈52%)	150
NNA-count	57 (≈38%)	26 (≈17%)	67 (≈45%)	150
NNA-mass	37 (≈25%)	23 (≈15%)	90 (≈60%)	150

Table 1: Some results of Experiment 2 by Braga et al. (2010).

A bivariate (*chi square*) test found significant differences between the mass vs. count as well as between the NA vs. NNA conditions: the non-prototypical bare mass nouns were more acceptable than the non-prototypical bare count nouns as subjects of individuating predicates while NA nouns were more acceptable than NNA ones. Braga et al. take this as evidence against Munn & Schmitt’s claim that mass nouns are ungrammatical with individuating predicates, therefore this argument for a different syntactic and semantic analysis of BSs and bare mass nouns fails. They suggest that the debate should be reopened and that natural atomicity rather than the mass/count distinction could be decisive for individuating predicates.

The striking fact about these results, as Braga et al. (2010: 85) themselves observe, is that neither of their conditions is very well accepted at all. The “best” result is obtained by NA count nouns (examples like (3)), but notice that there are still as many “acceptable” as “unacceptable” judgments. Neither Schmitt & Munn nor Braga et al. consider such sentences as ungrammatical and I am not aware of any paper containing such a claim. If we want to use AJTs in order to settle controversies of (un)grammaticality for BrP BSs, this remarkable behavior should be kept in mind and taken into account during the interpretation of the results. I will now turn to two experiments where the BSs again show such a behavior and then discuss all the results and interpretations of them, which will turn out to be problematic if the peculiar behavior of the BS is simply ignored.

5.1.2 Empirical controversy 2: Do BrP BSs denote kinds?

An alternative account of BrP BNs is provided by Müller (2000, 2002, et seq.), where BrP BSs as in (1) or (5) are analyzed as indefinite NPs under the scope of a generic operator. Thus, in her theory, (1), here repeated as (7a), is analyzed as (7b):

- (7) a. Professor trabalha muito.
 b. GEN[x:] (x is-teacher; x works-a-lot)
 Paraphrase: Usually, if x is a teacher, x works a lot.

Müller claims that this analysis can be generalized to all instances of BrP BSs. Crucially (for this claim), she treats certain instances of BSs which are accepted by other authors as ungrammatical. The cases in point are BS as subjects of predicates which select for kinds, and episodic predicates. (8) and (9) exemplify such sentences:

- (8) No ano 2030, gavião-royal vai estar extinto.
 in-the year 2030, hawk-royal will be extinct
 ‘In the year 2030, royal hawks will be extinct.’
 (Munn & Schmitt 2005:824)
- (9) [Até o século XX] Mulher vestiu saia.
 Until the century XX woman dressed skirt
 Women wore skirts until the 20th century.
 (Pires de Oliveira & Mariano 2011: 3748)

Simplifying somewhat, the so-called “kind predicates” are problematic because they reject indefinite NPs as subjects (cf. Krifka et al. 1995: 10ff): In the English translation of (8), the bare plural could be substituted by a definite singular NP but not by an indefinite singular one. Therefore, an analysis assuming inherently indefinite BSs would not work here. Episodic predicates in turn are problematic because they are not necessarily habitual or characterizing and therefore the generic operator would yield the wrong readings. This is not a problem for Müller, since such sentences are ungrammatical for her anyway. However, other researchers claim that such sentences are completely acceptable and should be treated as grammatical (Munn & Schmitt (2005), Dobrovie-Sorin & Pires de Oliveira (2008), Pires de Oliveira & Mariano (2011)). Such a situation obviously calls for an empirical test. For kind predicates, two AJTs are reported in the literature. As in the case of the BS vs. bare mass experiments, the ratings for the relevant items are not clear-cut in these two experiments, and interestingly, the authors come to opposing conclusions: Pires de Oliveira et al. (2010) claim that there is evidence for the acceptability of kind readings while Ionin et al. (2011) interpret their data as not containing such evidence. Both experiments are presented in the next section.

5.1.2.1 Testing Kind Predicates: Pires de Oliveira et al. vs. Ionin et al.

Pires de Oliveira et al. (2010) used the same experimental procedure as Braga et al. (2010). Episodic predicates were also tested in this experiment, but they will not be discussed later on. The experiment had four conditions (2x2 design):

- definite NP + kind predicate
- definite NP + episodic predicate
- BS + kind predicate
- BS + episodic predicate

For each condition, 6 items were created and the resulting 24 sentences were distributed on four lists together with 12 distracters, two of them being “control sentences” (one clearly grammatical and one clearly ungrammatical). Each list was judged by 50 participants. (10) is one of their “kind” items which was presented either with or without the definite article and (11) an “episodic” item for which the same manipulation was used. It was ensured that no item appeared both with and without article on the same list:

- (10) O Carlos leu que (a) baleia está em extinção.
The Carlos read that (the) whale is in extinction
'Carlos read that whales are on the verge of extinction.'
- (11) A Maria escutou que (o) carro entrou no Brasil no início
The Maria heard that the car entered in+the Brasil in+the beginning
do século.
of+the century
'Mary heard that the car was introduced to Brasil in the beginning of the century.'

In their results, Pires de Oliveira et al. do not give exact numbers for the conditions. They only provide bar charts, from which the figures presented in Table 2 were read off. They also report the approximate results for the control sentence considered as ungrammatical. I include it in Table 2 in order to show the different patterns. Table 2 is to be read analogously to Table 1.

Condition	“acceptable”	“undecided”	“unacceptable”	total
control (*)	≈10 (5%)	≈20 (10%)	> 160 (80%)	200
def. NP+kind	≈160 (64%)	≈40 (16%)	≈50 (20%)	250
BS+kind	≈120 (48%)	≈40 (16%)	≈90 (36%)	250
def. NP+epis.	≈160 (64%)	≈30 (12%)	≈60 (24%)	250
BS+epis.	≈80 (32%)	≈50 (20%)	≈120 (48%)	250

Table 2: Some results from Pires de Oliveira et al. (2010).

Pires de Oliveira et al. do not support their analysis with statistical tests, instead they look for patterns in the distribution of the raw figures. Given that there were more “acceptable” than “unacceptable” results for the BS+kind predicate condition, they conclude that “it is possible to claim that at least for some speakers, the sentences with BSs can be combined with kind predicates. Therefore, these sentences are not ungrammatical” (Pires de Oliveira et al. 2010: 132, my translation).

The experiment reported by Ionin et al. (2011) is especially interesting since it was conducted in three languages: English, Spanish and BrP. It is far more sophisticated than the previously reported one and I will only summarize the most important details for the points I want to make here concerning BrP kind predicates and anaphoric singular contexts. This experiment was conducted online (survey gizmo tool) and collected acceptability judgments on a scale from 1 (unacceptable) to 4 (acceptable). Five target sentences were presented simultaneously for rating after a paragraph-long context; the five sentences only differed in the form of the subject NP. The AJT consisted of 40 items: 20 target items and 20 fillers. The target items were broken down into five categories (four items each), testing canonical definiteness contexts for singular and plural in the control categories and what they call “NP-level genericity” (kind predicates) and “sentence-level genericity” (generic characterizing sentences) as well as mass nouns in generic contexts in the test categories. Ionin et al. do not report and discuss the mass category in their paper. The different NP forms tested were definite singular, indefinite singular, bare singular, definite plural and bare plural. The items as well as the order of the five simultaneously presented sentences were randomized by the software. (12) is an example item of a “definite singular” context followed by the five different NPs, (13) illustrates a kind predicate (for the sake of simplicity, the English version is given, the BrP sentences are direct translations):

- (12) Anaphoric singular context: Ralph has three pets: one dog and two birds. Ralph’s pets have very unusual habits. For instance. . .
- | | |
|--------------------------|-----------------------|
| a. The dog climbs trees. | (definite singular) |
| b. A dog climbs trees. | (indefinite singular) |
| c. Dog climbs trees. | (bare singular) |
| d. The dogs climb trees. | (definite plural) |
| e. Dogs climb trees. | (bare plural) |
- (13) NP-level genericity (kind predicate): I really like going to the zoo. Unfortunately, there are many animals that can’t be found in a zoo, or anywhere else. It’s very sad. For example. . .
- | | |
|--------------------------------|-----------------------|
| a. The dodo bird is extinct. | (definite singular) |
| b. A dodo bird is extinct. | (indefinite singular) |
| c. Dodo bird is extinct. | (bare singular) |
| d. The dodo birds are extinct. | (definite plural) |
| e. Dodo birds are extinct. | (bare plural) |

The prediction for kind predicates in English as well as in BrP would of course be that the definite singular and the bare plural should be OK in any case while for BrP the open question is the behavior of the BS. For the anaphoric singular control category, Ionin et al. assume that only the definite singular should be acceptable. Table 3 shows their results for the two above-mentioned categories.

	Brazilian Portuguese		English		Spanish	
NP-type	anaph.sg.	kind pr.	anaph.sg.	kind pr.	anaph.sg.	kind pr.
def.sg.	3,88	3,47	3,90	3,55	3,80	2,52
indef.sg.	1,88	1,26	2,09	1,47	2,14	3,58
bare sg.	1,84	2,34	1,27	1,32	1,13	1,36
def. pl.	1,62	3,88	1,53	2,26	2,06	3,89
bare pl.	1,83	3,80	1,75	3,96	1,38	1,70

Table 3: Some results (mean judgments) from Ionin et al. (2011).

Repeated measure ANOVAs were performed for the control and the target categories separately (crossing “context” (2 levels) with “sentence type” (5 levels)). All interactions were found to be statistically significant. Nonetheless, the authors also looked more closely at individual ratings since group results could hide possible variability there. Especially the BS+kind condition is difficult to interpret (again) because the mean rating lies almost exactly on the midpoint of the scale. As it is crucial for the following discussion, a lengthy quotation is in order:

“The more controversial part [...] is the obligatoriness of definiteness marking with singular kind terms. Specifically, we predicted that BrP speakers should disallow bare singulars in the context of NP-level genericity. In order to test this prediction, we checked whether the BrP participants rated definite singulars at least 0.5 higher than bare singulars in the context of NP-level genericity; this was indeed the case for 15 out of the 19 BrP participants (and indeed, in 14 out of these 15 cases, the rating difference between definite and bare singular NPs in this category was at least 1 point). Thus, definiteness marking does appear to be obligatory with kind reference, for most participants.” (Ionin et al. 2011: 980)

Consequently, Ionin et al. (2011: 984) conclude that their “current findings do not support the existence of kind readings of bare singulars”.

I would like to add a few words about the “anaphoric/definite” condition in Ionin et al.’s experiment because this condition is one instance of the definite/specific uses that will be tested in the experiments reported in section 5.2. Ionin et al. (2011:978) observe that the anaphoric BS condition was rated slightly higher in BrP than in the other tested languages (cf. Table 3), where its rating always stayed “at floor”. In their experiment, the BrP BS was also at floor in the anaphoric plural condition and the pairwise comparison showed that the better rating of the anaphoric singular is statistically significant. Nonetheless, since the other three NP types were rated equally high (cf. Table 3) and the BS was rated significantly lower than the other NP types in the anaphoric plural condition, they conclude that “consistent with the literature, bare singulars in BrP cannot be used in

anaphoric contexts” (Ionin et al. 2011: 978). In the overall discussion (5.3.1) I will argue against such an interpretation and support my position with the experimental results from Section 5.2.

5.1.2.2 Interim summary: gradient judgments in previous experiments

It is interesting and revealing to compare the method of data collection and interpretation in the three papers under discussion. As has been described above, Braga et al. (2010) and Pires de Oliveira et al. (2010) reduce the gradient judgments they collected into a rather binary pattern for the analysis: the results are aggregated into three groups and then compared to a binary pattern. The fact that the kind predicate sentences do not pattern with either of the “control” sentences provokes a different strategy of interpreting the results, namely generalizing the grammaticality claim not to the whole population but to a part of it (which is not specified).

Ionin et al. (2011) use the full four-point scale in their statistical analysis. However, given that all conditions and interactions are statistically significant, they recur to the *ad hoc* criterion of interpreting a distance of 0.5 points on the scale as meaningful for differences of grammaticality, without justifying this measure.

For now, the most important observations are (1) that the aim in all these papers is to separate the grammatical combinations from the ungrammatical ones, applying a binary concept of grammaticality directly to the data and (2) because the ratings constantly concentrate somewhere in the middle of the used scales or are balanced in the extremes, the authors resort to different alternative strategies in applying the binary approach. This leads to opposing conclusions based on quite similar data.

It is far beyond the scope of this paper to solve the problem of interpreting intermediate acceptability or grammaticality judgments in general. A first step in this direction should be to find a consensus with respect to what these judgments are and how they come about. A point of departure for such a consensus could be Bader & Häussler’s (2010) account based on Signal Detection Theory which I will introduce briefly in the following section and subsequently also explore in the discussion of the experiments.

5.1.3 A model of gradient and binary perception of well-formedness

Bader & Häussler (2010) propose a theory of grammaticality judgments based on Signal Detection Theory (=SDT), “a general theory of how stimuli are classified into discrete categories” which “gives a high-level characterization of the mental processes that are responsible for converting a gradient percept of sentence acceptability into an overt binary grammaticality judgment” (Bader & Häussler 2010: 306). It is a particularly interesting approach, first, because it departs from a

more general theory and second because in this model, the gradient perception of well-formedness is prior and the basis also in the case of binary judgments⁶:

“In order to apply the theory, we decompose the process of giving a binary grammatical judgment into two major steps. During the first step, each sentence is assigned a continuous acceptability value. The second step maps this continuous value onto a binary classification, which in turn determines the overt response.” Bader & Häussler (2010: 307)

Bader & Häussler (2010: 310) admit that at the moment, their model is not a real processing model but rather an abstract mathematical one, with their first step being nothing more than a black box. They leave the question of how the acceptability scores are actually computed for future studies and for the time being just hypothesize that the results from magnitude estimation experiments map onto the continuous acceptability scores which are the output of the black box and that the latter can be estimated from the former. Still, their conclusion that the underlying judgment is inherently gradient deserves emphasis.

Although leaving many questions open about the “black box”, Bader & Häussler mention some possible external influences and problematic mismatches between them: They discuss the factor frequency and come to the conclusion that “for the syntactic constructions considered in this paper, the relationship between perceived well-formedness and frequency of occurrence does not seem to be a random one” (Bader & Häussler 2010: 315). The mismatch arises when “speakers deny that they use or even know a certain structure when explicitly queried, but then use the structure nevertheless (see for example Labov 1977)” (Bader & Häussler 2010: 316). The authors leave it open whether such cases are in fact real exceptions with regard to the influence of frequency, since their experiments do not contain structures of this type. The experiments on BrP BNs reported in this paper later on, however, contain exactly this kind of structures. Of course the design of the experiments was not developed in order to answer Bader & Häussler’s question, but the data can nevertheless be explored in this direction. At least for BrP BNs I think it can be tentatively said that these cases are no real exceptions. In controlled experimental contexts, averaging over speakers, it is the case that structures which are in fact used by speakers (even the marginal ones) can be shown to be rated higher than forms that just do not exist – if the judgments are collected with the necessary degree of fine-grainedness.

One such phenomenon, where apparently speakers reject actually attested forms quite robustly is the case of definite/specific uses of BSs in BrP. This phenomenon will be introduced in the next section, followed by the results of two AJTs and a production test on this phenomenon reported in section 5.2.

⁶ Bader & Häussler develop their theory by testing and comparing different judgment methods, namely magnitude estimation, speeded binary grammaticality judgments and “off-line” binary grammaticality judgments. The phenomena tested are German word order, case and argument alternation.

5.1.4 The case of definite/specific bare singulars in Brazilian Portuguese

Outside the formal semantic/generative literature cited above, sporadically one finds the claim that beside the other contexts in which BSs occur, they can also be used referentially in definite and/or specific contexts (cf. Amaral (1920), Kabatek (2002)). Examples of such cases are (14) and (15):

- (14) o carro do moleque num tinha documento
 The car of+the boy not had document
 [...] **documento tava na chácara**
 document was in+the cottage.
 ‘[...] The registration document was in the cottage.’ (IBORUNA)
- (15) Nariz já, mas supercílio não. **Nariz já saiu muito sangue.**
 nose already but eyebrow not nose already went-out much blood
 ‘The nose, not the eyebrow. **The nose has been bleeding a lot.**’ (PEUL).

Such or similar examples are widely ignored in the formal/generative literature. This might be due to the fact that they are considered ungrammatical by almost everybody: “[...] the bare singular is not allowed in subject position when the interpretation of the NP is specific, designating just one particular individual” (Santolin 2006: 27, my translation). For indefinite contexts, usually only “narrow scope” (opaque) readings are considered as available (cf. Schmitt & Munn 2003, Müller & Oliveira 2004; among others). This rules out specificity. And for definite NPs, Dobrovie-Sorin & Pires de Oliveira (2008: 112) claim: “in all the languages that have an overt definite article, the iota must be overtly realized as a definite article”. Pires de Oliveira et al. (2010) and Ionin (2011) predict that due to ungrammaticality, these sentences should receive low ratings in their experiments.

However, there is good reason to treat the introspective judgments with special care since the experimentally collected judgments described above show that there are strong discrepancies, and these are, it seems to me, just the tip of the iceberg⁷. Therefore, before further generalized claims about grammaticality can be made, it is necessary to clarify the status of these definite/specific BSs. By saying that the reported intuitions on BrP BNs must be treated with special care I am not suggesting that they are completely untrustworthy. Of course the fact that many speakers reject sentences like (14) or (15) also calls for an explanation. And in fact, the use of BNs in these contexts is clearly rather marginal and the definite article is predominant in such NPs. This is also reflected in the judgments presented in this paper. What I want to question is that the rejection can simply be attributed to ungrammaticality. Ungrammaticality in the strict sense would mean that the BSs are

⁷ See e.g. the special issue on BNs of the *Journal of Portuguese Linguistics*, no. 09, 2010 featuring many examples marked with a question mark or reports of diverging judgments by different speakers.

banned in certain structures. This implies that the only way to maintain the ungrammaticality claim for definite/specific contexts in the face of examples like (14) and (15) is to account for them on extra-grammatical grounds, e.g. by performance factors. However, it is not clear what such an account could look like and whether it would really be desirable. Rather, I would like to consider a recent approach by Pires de Oliveira & Rothstein (2011, in press) which suggests that there are no syntactic restrictions on BrP BSs at all.

Pires de Oliveira & Rothstein (2011, in press) treat BrP BSs as mass terms denoting kinds. They use data analogous to those from Braga et al. (2010) in order to show that the arguments against treating BSs as mass terms are not valid since they only work for the prototypical mass and count nouns. Given the fact that for many speakers BSs are fine with kind predicates (Pires de Oliveira et al. 2010) and that even in episodic sentences BSs may occur freely if the context of the sentence supports a kind-reading for the BS (cf. example 9 above), the authors discard analyses based on indefiniteness. They suggest that the observed restrictions are better captured by the semantic and pragmatic mismatches between the kind-denotation of the BS and the respective context (of the sentence as well as of the whole utterance). In Pires de Oliveira & Rothstein (in press) they also sketch an analysis for examples like (14) & (15), in which the definite/specific reading arises through accommodation. The definite/specific reading is pragmatically “enriched” upon the kind semantics through the anaphoric use of the BS and the fact that only one possible referent is available in the context. A more precise characterization of the behavior of such anaphoric BSs is yet to be given. Natural starting points would be some of the questions regarding anaphora resolution which are discussed in this volume (especially chapter 4, 6 and 7), such as for instance the cross-linguistic stability of certain pragmatic principles.

Rather than delving deeper into the details of this approach or alternative explanations, the main goal here is to argue against the apparent ungrammaticality of certain instances of BSs and to suggest that the whole debate about judgments of (un)grammaticality in this case needs a better theoretical foundation. Without that the peculiar behavior of BSs in AJTs makes an objective interpretation of such tests impossible for many theoretically important configurations. From this perspective, Pires de Oliveira & Rothstein’s claim with regard to the absence of any syntactic restrictions on BSs in BrP is quite appealing. But before I try to tie up all the loose ends in the discussion, I will present additional evidence for my position from three experiments.

5.2 Three experiments on specific/definite BNs

The first two experiments presented here were designed to test the authentic but marginal sentences found in corpora of written and spoken language. The third one aimed at eliciting the production of definite/specific BSs. Since there is hardly

any experimental work on BrP BSs, especially on the definite/specific ones, the experiments have an exploratory component: How do such BSs behave in different AJTs? The main hypothesis behind Experiment 1 and 2 is that contrary to what most of the literature says, speakers would not rule them out categorically. The expectation was that specific/definite BSs would behave like the other BSs. For Experiment 3, the hypothesis was that if specific/definite BSs are in fact grammatical, it should be possible to stimulate their production.

Experiment 1 used a sample of original sentences found in corpora of written and spoken language (henceforth: “attested sentences”) featuring BSs. These sentences were contrasted with ones that differ only in including the definite (and where possible also the indefinite) article, which would be the canonical form.

In **Experiment 2**, specific/definite BNs were tested against a range of filler items with different degrees of well-formedness, thus fully exploring the idea of the gradient perception of well-formedness. The experiment made use of the “Thermometer Judgment” method (Featherston 2008).

Experiment 3 was an elicitation task testing the spontaneous production of BNs, especially whether their frequency can be increased through previous exposure to similar structures.

For purposes of motivation, in the case of Experiment 1 & 3, participants were offered a reward for participation, in the case of Experiment 2 there was a lottery with three prizes. However, almost half of the participants preferred (and insisted on) participating entirely voluntarily⁸.

5.2.1 Experiment 1: 5-Step offline questionnaire

Experiment 1 was conducted as a pencil & paper task in São Paulo and Rio de Janeiro. The participants performed the tasks either individually or in small groups of maximally five persons. There was no communication between participants in the latter cases, the experimenter was always present.

Participants: 30 Participants (origin: São Paulo (15), Rio de Janeiro (9), 6 part. from three other states); age: 19-31 (mean: 22,7; standard dev.: 3,2); gender: 20 fem., 9 male; 1 no inf.; education: undergraduate and graduate students; subjects studied: “Letras”⁹ (24), other (2), no inf. (4).

⁸ All the sentence materials, instructions and a digital version of the questionnaire are freely available on-line:

Experiment 01: <http://hdl.handle.net/11022/0000-0000-1CCA-E>

Experiment 02: <http://hdl.handle.net/11022/0000-0000-1CCB-D>

Experiment 03: <http://hdl.handle.net/11022/0000-0000-1CCC-C>

⁹ The “Letras” curriculum includes courses of language, literature and linguistics. Proficiency in second languages was not tested. It was also not asked explicit-

Method & material: An 18-page questionnaire booklet containing the experimental material was distributed among the participants. Additionally, participants received written instructions for each of the five tasks and an example for the rating task (3). They were asked to keep the instructions on the table beside the booklet and to consult them as often as necessary. They were also allowed to ask the experimenter in cases of doubt before the experiment started. There was no time limit to complete the questionnaire. On average, the experiment took between 40 and 60 minutes.

The materials of Experiment 1 consisted of attested sentences containing a BN in argument position and slightly modified versions of these sentences. The original sentences came from different sources and had different readings, as shown in Table 4. Three of the tested items are not included in Table 4 because they turned out to be otherwise marked and therefore had to be excluded from the analysis. The non-specific items from written and spoken language are presented together because they should not be considered marginal and problematic and will not be discussed here in detail.

Source:	spoken		written		sp. & wr.
Reading:	def./spec.	ind./spec.	def./spec.	ind./spec.	non-spec.
Nr./items:	5	3	10	6	9

Table 4: Items and groups for Experiment 1.

To each of the 36 original BN-sentences I added a counterpart with a definite article, and, in the case of the indefinite groups and the non-specific group, one with an indefinite article as well. The resulting two or three sentences were presented in a block, one after the other (resulting in a total of 36 blocks/92 sentences). The order of the sentence blocks was randomized but for practical reasons all participants saw the same random order.

Design: For each block, the participants had to complete five tasks. The English equivalent of the instructions for each step is given below together with a short explanation of the function of each of them.

- Task 1: Guess which kind of text the original example might be taken from.
- Task 2: Judge the sentences on a scale from 1 (not natural) to 7 (natural).
- Task 3: Indicate differences in meaning between the sentences (if applicable).
- Task 4: Indicate impossible sentences (none, one or more).
- Task 5: Guess which of the sentences might be the original one.

ly whether the informants grew up monolingually or not but for most of them this can be safely assumed.

The first step was necessary, since the sentence material contained examples from very different sources which were presented out of the blue¹⁰. As different types of text might have different restrictions on BNs, this task forced the informants to contextualize the sentence before judging it. The most important distinction was between the more formal/written register vs. spontaneous spoken language. Step 2 is of course the heart of the experiment and its function is self-evident. In step 3 participants could give additional information that seemed relevant to them but could not be expressed in the judgment task. It also served to separate step 2 from step 4 in order to avoid a consecutive execution of the two judgment tasks. Step 4 was intended to simulate a binary (grammaticality) judgment after the gradual one. Since the participants knew from the instructions that one of the two or three sentences was the original, attested one, step 5 was included in order to give them another opportunity to provide additional information which might be interesting for the interpretation of their performance. It also served as a bit of distraction and change of focus away from sentence judgment before going over to a new sentence block.

Results: For this experiment, 30 data sets were collected; none of the subjects was excluded from the analysis. There are occasionally a few missing data points, which, however, were not considered as decisive for excluding the whole dataset. In sum, less than 1.7% of the possible data points are missing; the maximum is five per question and subject in Task 1, and one single case with three missing data points in Task 2.

Task 1: In 89.2% of the cases (standard dev. 10.1), participants suggested the correct original source of the example (formal/written register vs. spontaneous spoken language).

Task 2: Table 5 reports the mean judgments for the gradient ratings. In order to have a symmetrical design for the statistical analysis, two groups of items were excluded, namely the items containing definite articles in the indefinite groups (spoken & written). The non-specific items were not considered, either.

Source:	spoken (sp.)		written (wr.)		sp. & wr.
Reading:	def./spec.	ind./spec.	def./spec.	ind./spec.	non-spec.
BS	3.43	2.68	3.12	4.61	4.71
def NP	6.57	6.53	6.29	5.61	6.16
ind NP	-	5.26	-	6.22	5.25

Table 5: Mean gradient judgments from Experiment 1.

¹⁰ Sometimes, if the interpretation of the target sentence without context was difficult or unclear, the previous sentence from the text was included (9 cases). In this case, the target sentence was presented in bold letters and the instruction was to judge only the sentence in bold.

Repeated measures ANOVAs revealed that two of three main effects were strongly significant ((i) & (iii)), (ii) only marginally: (i) written vs. spoken $F(1,29)=33.081$, $p<0.001$; (ii) def. vs. indef. $F(1,29)=2.007$, $p=0.167$; and “article+N” vs. BS $F(1,29)=147.854$, $p<0.001$. Furthermore, all interactions were significant: written/spoken X def./ind. $F=103.055$; $p<0.001$; written/spoken X BS/article $F=5.563$; $p=0.025$; def./ind. X BS/article $F=36.123$; $p<0.001$.

Since the sentence material in the different groups was very different, ANOVAs were also performed for all “minimal pairs” of “article+N” vs. BS. All were significant at $p<0.001$; SD: $F(1,29)=135.531$; SI: $F(1,29)=62.682$; WD: $F(1,29)=228.78$; WI $F(1,29)=42$.

Task 3: It goes beyond the scope of this paper to discuss the answers in Task 3.

Task 4: Table 6 contains the mean rejection rates for the 36 BSs and their counterparts with the appropriate article (definite/indefinite).

Source:	spoken		written		sp. & wr.
Reading:	def./spec.	ind./spec.	def./spec.	ind./spec.	non-spec.
BN:	42%	63%	42%	31%	23%
Art + NP	7%	12%	1,3%	1%	8%

Table 6: Mean rejection rates for the 5 target groups in Experiment 1.

In order to verify whether there is a correlation between gradient judgments and rejection rate, two linear regressions were calculated for the two values with the means of each example: One for the sentences containing articles in the target NP and one for the BS sentences. The regression lines are shown in Figure 1.

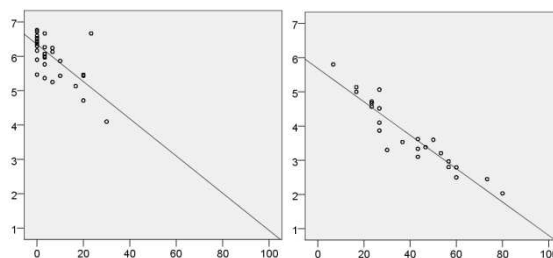


Figure 1: Linear regression for sentences with article+N (left) and BSs (right). For each item, mean gradient judgments (y-axis) are plotted against mean percentages of rejection (x-axis).

The coefficient of determination for the sentences with an article is $R^2=0.486$ and for the BS sentences $R^2=0.875$ ($R^2=0.878$ for both together). Thus, there is a strong linear correlation between gradient judgment and rejection rate.

Task 5: In Task 5 the participants had to guess which example in a given block might be the original one. Table 7 summarizes the answers suggesting the BN sentence.

Source:	spoken		written		sp. & wr.
Reading:	def./spec.	ind./spec.	def./spec.	ind./spec.	non-spec.
% for BN	9	4	7	19	35

Table 7: BN-examples considered as the original examples.

Summary of Experiment 1: The overwhelmingly correct assignment of the target sentences to the kind of text it was taken from suggests that Task 1 presented no problem for the participants. In particular, there was no difficulty to tell apart the written language examples from the spoken language examples. Since the main difference in BN uses in the experimental items lied in this differentiation, the function of Task 1 can be regarded as fulfilled and therefore interferences in judgments because of this kind of influence can be considered minimal.

The gradient ratings from Task 2 show significantly lower judgments for specific BSs in all contexts. The judgments for the BSs group slightly below the middle of the scale for the examples from spoken language, and around the middle for those from written language. In this paper, only the results from the pairwise comparisons are interpreted since in the other cases influences of the sentence material cannot be ruled out. In Task 4, the BS sentences were not rejected categorically. With the exception of the indefinite/specific sentences from spoken language, all other BSs had a rejection rate clearly below 50%. Moreover, a strong linear correlation was found between the mean gradient judgment and the mean rejection rate.

5.2.2 Experiment 2: Thermometer Judgments

Experiment 2 contrasts definite/specific BSs and their “article+N” counterparts with fillers which show a wide range of well-formedness. It also corrects shortcomings of the design of Experiment 1.

Participants: 48 participants (origin: Rio de Janeiro (26), São Paulo (12), 10 part. from five other states); age: 19-57 (mean: 27,8; standard dev.: 10,1); gender: 36 fem.; 12 male; education: high (univ. students or holding univ. degree); subjects studied: “Letras” (22, cf. fn 10); other (25), no inf. (1).

Method: Experiment 2 made use of the Thermometer Judgment paradigm (Featherston 2008). Thermometer Judgments are a special incarnation of the magnitude estimation method: The “naturalness” of sentences has to be rated with respect to two reference sentences which were given fix points on an (open-ended) numerical scale, the “not natural” sentence at 20 and the “completely natural” sen-

tence at 30. These two reference sentences and their respective values remain on the screen during the whole experiment.

In the instructions, the concept of magnitude estimation was explained and exemplified first with lines of different length and then with sentences. Before the actual experimental materials were shown, there were two practice phases, one with lines of different length and one with sentences. As described above, in Experiment 2, all sentences were introduced by a context sentence. The context sentence fixed the intended reading of the BN in the target sentence. Context and target sentence were presented together below the reference sentences, the target sentence was presented in red letters while the others appeared in black letters. The English equivalent of the task question was “With respect to the two target sentences, how natural is this sentence?” The experiment was conducted via internet using the WebExp2¹¹ software.

Materials & Design: The materials consisted of 32 items and 12 fillers whose presentation was randomized by the software. The only manipulation in the target sentences was the presence vs. absence of the determiner in the target NP. The sentence materials were largely inspired by the attested example sentences used for Experiment 1 while controlling for the factors listed in Table 8¹²:

	Target items: 32	Manipulation	
Contr. factors / “Conditions”	SYN. pos. of Target NP & DEFiniteness	NP (type)	Noun type ¹³
1	preverbal subject/ definite (8)	Def. art. +N	concrete (4)/ abstract (4)
2		BS	
3	Preverbal subject/ Indefinite (8)	Indef. art. +N	concrete (4)/ abstract (4)
4		BS	

¹¹ <<https://wiki.inf.ed.ac.uk/WebExp/WebHome>>

¹² The 12 filler items (controlled factors 9-17) featured 5 well-formed sentences: one not stigmatized colloquial construction (9), three unmarked sentences (10), one generically interpretable sentence (11); and 7 not well-formed ones: two with wrong prepositions (12), one stigmatized coll. constr. (13), one with agreement error (14), one highly stigmatized coll. constr. (15), one with a semantic mismatch (16), and one with a syntactic and semantic mismatch (17).

¹³ The sentences were balanced for concrete and abstract nouns within each of the “conditions” (1-8). For each of these conditions, in one of the sentences the NP consisted of the noun only, while the other three were modified: one by and adjective, one by the possessive “dele/dela” (a contraction of the preposition *de* and the pronoun *ele*) and one by a full PP. The intention behind this manipulation was that these structures were frequent in the corpus examples and I wanted to see whether modified BNs are rated considerably different from the unmodified ones. These systematic differences however were not treated as “real” conditions in the experiment since that would have proliferated the number of items immensely.

5	post-verbal object/ definite (8)	Def. art. +N	concrete (4)/
6		BS	abstract (4)
7	Post-verbal object/ Indefinite (8)	Indef. art. +N	concrete (4)/
8		BS	abstract (4)

Table 8: Distribution of target items and conditions in Experiment 2.

To make sure participants didn't see both versions, the items were divided into two lists of 44 items. Each participant judged the items of only one list. No. 1-8 from Table 8 were treated as "conditions" in the sense that groups according to the syntactic position and the lexical semantics of the noun could be put together for the analysis. The context sentence was used to ensure the intended reading of the NP (definite or indefinite). For the definite reading, this was done by either providing a unique referent of the NP, e.g. by introducing it in the first sentence or by describing a situation where the uniqueness of the referent is guaranteed (cf. (16)). For the indefinite reading, it was ensured that no such "uniqueness presupposition" was triggered (cf. (17)).

- (16) No final do jogo, Carlos tem que bater o penalty decisivo.
in+the end of+the game Carlos has that hit the penalty decisive

O coração dele está batendo forte.

the heart of+him is beating strong

'In the end of the game, Carlos has to shoot the decisive penalty. His heart is beating heavily.'

- (17) João não consegue dormir por causa do barulho lá fora.
João not succeeds sleep for reason of+the noise there outside

Um cachorro está latindo na rua.

a dog is barking in+the street

'João has troubles with falling asleep because of the noise outside. A dog is barking on the street.'

Results: For the analysis, judgments were transformed into z-scores in order to balance out extreme judgments. Figure 2 shows the mean z-scores and standard error bars for all eight conditions (1-8) and all types of fillers (9-17). The horizontal line at z-value -0,5 separates those filler items considered grammatical from the ungrammatical ones.

Repeated measures ANOVAs were performed for the eight target conditions. There was one significant main effect (NP) and of the interactions only one was significant (SYN X NP)¹⁴. Due to considerable differences in the sentence materi-

¹⁴ Not significant main effects: SYN: $p > 0.2$; DEF: $p > 0.2$; Sign. main effect: NP: $F(1,47)=113.588$; $p < 0.001$. Significant interaction: SYN X NP: $F(1,47)=5.312$, $p=0.026$. The other interactions all had $p > 0.6$.

als, repeated measures ANOVAs were conducted for all “minimal pairs” of (in)def. art+N vs. BS (conditions 1-8, the BS condition in each case is the lower rated one): All of them proved significant at $p < 0.001$ (1 vs. 2: $F(1,47)=49.861$; 3 vs. 4: $F(1,47)=75.185$; c. 5 vs. 6: $F(1,47)=48.653$; c. 7 vs. 8: $F(1,47)=53.129$).

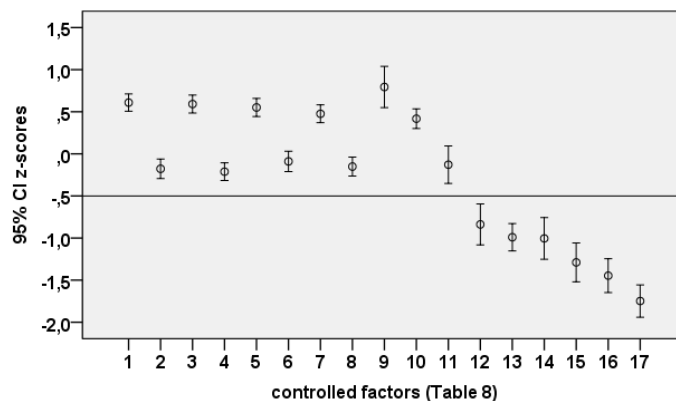


Figure 2: Z-scores for conditions 1-8 and fillers from Experiment 2.

Summary of Experiment 2: The only clear effect in this experiment resulted from the determiner manipulation. None of the other manipulations (definite/indefinite, concrete/abstract, syntactic position (subj/obj), or the four kinds of NP (modified or not)) led to statistically significant differences. But since the sentence materials differed across conditions, this lack of effect cannot be interpreted and the question must be left open for a follow-up study.

What becomes immediately visible is that the definite/specific BSs are in fact rated considerably lower than their “article+N” counterparts (replicating the results of Experiment 1). On the other hand, they are judged better than any fillers showing some violation of well-formedness.

5.2.3 Experiment 3: Elicitation

Experiment 3 was conducted in São Paulo, Rio de Janeiro and Niterói. Its aim was to investigate the distribution of BSs in elicitation tasks and to see whether the frequency of BS occurrences could be increased.

Participants: 36 participants (origin: Rio de Janeiro (17), São Paulo (14), 5 part. from three other states); age: 18-51 (mean: 27,3; standard dev.: 6,9); gender: 22 fem.; 14 male; education: 13 “high” (university stud.; 11 from “Letras”, cf. fn 10; 2 others); 7 “middle”; 16 no inf.

Methods & materials: Participants had to read a short summary of the plot and then narrate a picture story. Two different kinds of summaries were used: One group of participants, serving as a control group, received summaries in which the plot was laid out in short and normal well-formed sentences. The second group received summaries in a “telegraphic style”: Instead of short sentences, the summaries imitated telegraphic messages, where most functional elements, e.g. conjunctions, and all articles were eliminated. Both groups were told that the summary would give them a basic idea of what the story was about and thereby would help them to tell a story they have probably never seen before. Four picture stories (from Mercer Mayer’s frog stories¹⁵) served as stimulus materials. Participants read each summary and re-told the story immediately afterwards. The pictures were shown on a laptop screen and the narration was recorded with a ZOOM H2 recorder in a closed room. Beside the narrator and the experimenter, no other person was in the room. (18) and (19) give the first sentences of a “normal” and a “telegraphic” summary:

- (18) O menino e o cachorro têm como animal de estimação um sapo. O sapo vive dentro de um vidro...
 ‘The boy and the dog keep a frog as their pet. The frog lives in a jar...’
- (19) a noite, menino e cachorro olham para sapo dentro de vidro - menino e cachorro dormem - sapo pula de vidro...
 ‘in night, boy and dog look for frog in jar – boy and dog sleep – frog jumps out of glass...’

Before the results are presented, a few words are necessary about the somewhat unconventional design. The idea behind the different summaries was that previous exposure to many “BSs” could also increase the frequency of BS production. Group 1 served as the “control” group, for comparison with the “primed” Group 2, which read the telegraphic summaries. Since a summary consisting of well-formed sentences should not have any direct influence on the production of BSs, the behavior of Group 1 is considered “normal”. Of course the BSs presented in the “telegraphic” stimulus material and the BSs expected in the production task are of a very different kind. Obviously, the aim was not to elicit telegraphic narrations. However, in the first test runs most of the narrations were quite natural, thus the BSs produced in fluid narrations of Group2 were taken as valid occurrences. The few nonfluid narrations were excluded from the analysis (column “T” in the results table). Any narration from Group 2 containing a sentence-initial BS in two subsequent sentences was considered as telegraphic and the whole narration was excluded. If this occurred in one of the first three stories, all the other stories from

¹⁵ I did not ask explicitly whether the participants were familiar with the stories. Based on their reactions, however, I strongly assume that none of them were. I thank a reviewer for bringing up this point.

this participant were excluded as well. This is a very rigid criterion, but I wanted to exclude telegraphic passages as effectively as possible. Furthermore, the goal was not simply to increase the absolute frequency of BSs but rather to see whether they would be produced equally in different syntactic positions. Therefore, about 50% of BSs in the telegraphic summary would have been subjects in full sentences. The other 50% would have been objects or were inside PPs.

Results: Narrations had an approximate duration between 2 minutes and 8 minutes, most of them around 5 minutes, thus the total recordings per participant range from about 10 minutes to 27 minutes, most of them around 20 minutes. The total recorded material is 8h 56min. The stories were transcribed by me in a semi-orthographic way paying special attention to the presence or absence of the definite article. For all critical cases, spectrograms were used as decision guidance. Afterwards, all relevant noun phrases were extracted for analysis. The results are summarized in Table 9:

	Group 1	Group 2	T
Stories	76	55	12
Relevant NPs	3606	2727	597
Not analyzable ¹⁶	613 (17 %)	474 (17.4 %)	80 (13.4 %)
Prev.Subj. BNs	56 (1.6 %)	100 (3.7 %)	84*+54** (23 %)
Postv.DO BNs	1?	1?	1?
Subj. & DO BN	0	1	16 (2.7 %)
BNs inside PP	0	0	2

Table 9: Results from Experiment 3. (* BS+VP; ** BS+[VP+BS])

Of the 144 stories, 12 were excluded because they contained telegraphic passages (column “T”), one story was lost due to technical problems. The BS production in the remaining 131 stories of this experiment can be summarized as follows: Even in the control condition, about 1.6% of all relevant NPs which, according to the literature must occur with an article (definite/specific singulars) were produced clearly without one. This rate was more than doubled by previous exposure to BSs in the telegraphic style summary. This becomes even more interesting when we observe the third generalization: Although the total number of BSs could be increased, the distribution remained the same. The presentation of BSs in direct object position and inside PPs did not lead to an increase of BS uses by Group 2 in

¹⁶ Studies on the BrP article system and on BNs mostly ignore the problem that the definite article is just a vowel and BrP has strong phonological assimilation processes (“External Sandhis” (Bisol 2003)) that frequently affect the definite article as well. These processes are responsible for the uninterpretability of about 17% of all possible article contexts in this corpus. I think this fact must also be taken into account in the discussion on BrP BNs. But it goes beyond the scope of the present article to point out the implications that arise from these facts and this important question must be put aside for another occasion.

these contexts. As in the control condition, practically only preverbal subject BSs were produced in the primed group. The increase of BS through previous exposure is statistically significant: Relative frequency means for each subject were calculated¹⁷ and compared in a t-test. The t-test for independent groups gives the following results: $t(17,076)^{18} = -2.322$, $p=0.033$.

The frequency of 1.6% in the control group seems to be insignificantly small, so why bother about it? I think there are good reasons not to discard marginal data as negligible a priori and I do not agree that in such cases, first “the reader has [...] to be convinced of its theoretical importance”, as an anonymous reviewer puts it. Firstly, the narrations represent only a very special kind of language use and the (admittedly very marginal) occurrence cannot be generalized to the language in general. This is therefore to be considered as an initial step towards a more comprehensive corpus study. It should also be noted that the 17% of unanalyzable cases presumably contain more examples which could not be identified with confidence due to specific vocalic contexts. Thus, the 1.6% figure is the result of the most conservative analysis, the “100% waterproof” examples. More importantly, however, as was shown in section 5.1.2, other marginal constructions play an important role in the case of BrP BSs¹⁹ and surely not only there. If our theories are built on more and more fine-grained analyses, it is unreasonable to take disproportionally more coarse-grained criteria for relevance in the verification of the theories. This point will be taken up in the general discussion.

5.3 Discussion

To sum up, the main hypothesis for Exp. 1 & 2 was confirmed: Speakers did not rule out definite/specific sentences categorically, instead, the intermediate ratings are reminiscent of the general behavior of these forms in AJTs, as in the ones presented in 5.1.2; NPs preceded by articles usually are preferred. The results of the other factors in Exp. 1 & 2 seem to be contradictory (all comparisons significant vs. nothing significant), however, it must be kept in mind that these factors had a rather exploratory character and they used different sentence materials. This clearly deserves further and better controlled studies. Exp. 3 gives a first hint at the distribution of anaphoric BSs in language production. The results show that the sentence-initial/subject position is the strongly preferred one (maybe the only possible one) since the BSs almost exclusively occurred there and only in this position could frequency be increased through priming, although other possible syntactic positions were primed as well. This patterns nicely with findings from spo-

¹⁷ Raw frequency of BSs/6333 (sum of relevant NPs).

¹⁸ Equal variances not assumed.

¹⁹ I invite skeptics to have a look at frequency reports in the corpus study on BS+kind predicate sentences in Pires de Oliveira et al. (2010) if the figures presented here seem insignificant.

ken language corpora. Wall (in press) presents such data and argues that this distribution is not due to syntactic constraints but to the interpretations licensed by information structure. Interestingly, the importance of information structure for anaphoric expressions is also highlighted in chapter 6. While in both cases it seems to be only one of several factors that come into play, this completes the picture of cross-linguistically stable aspects in the interpretation of linguistic expressions a bit more.

What remains to be discussed is the status of definite/specific BSs, given that almost all studies treat them as ungrammatical, and the question what the peculiar behavior of BSs in ATJs could mean for linguistic theories. This will be done in the following two sub-sections which end with the suggestion that in the case of BrP BSs, but probably also for other marginal phenomena, it is necessary to combine different approaches along the lines sketched in this paper. Approaches based on one method are not sufficient.

5.3.1 The definite/specific BSs in the light of Bader & Häussler's theory

As mentioned in section 5.1.2.2, Ionin et al. (2011), interpreting the results of their AJTs, come to the conclusion that the BS “cannot be used in anaphoric contexts” in BrP, although it is judged better than in the other two languages and although this is the only condition where this NP type does not “stay at floor” (except for BrP generic sentences, of course). As pointed out at the end of section 5.1.2.1, I think there are good reasons to doubt their conclusion. Ionin et al.'s experiment is a very ambitious one testing five different NP types in five different contexts, resulting in a total of 25 conditions, with four observations per condition. Since using a standard Latin square design would not have been viable, some trade-off was necessary. The authors decided for a much simpler design – but at a very high cost: The distribution of items in different conditions on different lists was abandoned entirely. The target sentences with different NPs were presented **simultaneously** for evaluation. This must be kept in mind and in my opinion it makes the authors' interpretation of the judgments very dubious. I do not say that it is entirely impossible to interpret the reported judgments (my Exp. 1 uses a very similar procedure), but one should be aware of the kind of data one is dealing with in this special case. The authors do not seem a bit concerned about that, saying succinctly that “[t]he participants were explicitly instructed that they did not have to rank the sentences, and could give the same rating to two or more sentences” (Ionin et al. 2010: 973). However, even with the instructions just quoted this design invites to compare sentences, or at least makes it inevitable. Thus, it cannot be ruled out that clear prescriptive/normative pressure and best choice strategies could bias the results penalizing the BS. I would therefore predict that by correcting this drawback, the ratings for BSs should not be so drastically lower than the others, probably even exceeding the others (obviously except for the definite singular NP). The results of Exp. 2 support this view. There, the BS sentences were

not presented together with their “competitors” and did not receive such extremely low ratings. This criticism extends of course to Ionin et al.’s conclusions about BSs and “NP-level genericity” (kind predicates). Since in these cases the ratings of the BSs are much closer to their “competitors” than in the definite/specific context, there should be even less doubt about their grammaticality. In a sense, Exp. 1 replicates Ionin et al.’s findings that when direct comparison of “conditions” is possible, the ratings for the BS stay in the lower part of the scale however, not “at floor”. As in Ionin et al., all manipulations produced statistically significant effects, but by now it should be clear that this finding is difficult to interpret in terms of (un)grammaticality (recall that Ionin et al. recur to the ad hoc criterion of a difference of 0,5 points on their scale in the interpretation of the kind predicate condition). And it also should be clear that interpreting isolated experimental results on marginal phenomena is not without risk. Additional information is needed. The rejection task in Exp. 1 (the closest to what Bader & Häussler would call “grammaticality judgments”), for instance, shows that on average, the definite/specific BSs were rejected in less than 50% of the judgments. Arguing along the lines of Pires de Oliveira et al. (2010) (cf. 5.1.2.1), this can be taken as evidence that at least for a group of speakers (in certain circumstances), these forms are grammatical. But I think that there is no need for recurring to such alternative strategies since stronger claims can be made: If we compare these results with those of Exp. 2, where the different conditions could not be compared directly, the BS still was rated significantly lower than the article-headed NP, but compared to the filler items with different degrees of well-formedness, they tended to pattern with the grammatical items (cf. Bott & Schlotterbeck (2009) for a similar strategy). In terms of Bader & Häussler’s (2010) theory of how gradient perceptions of acceptability are mapped onto binary grammaticality judgments, one could interpret the results from Exp. 2 as showing that even definite/specific BSs are perceived as being better than the cut-off value for (un)grammaticality. The strong linear correlation between the gradient and binary judgments of Exp. 1 also suggests that cases where speakers declare forms which they actually use as ungrammatical should not be treated as exceptions to Bader & Häussler’s theory. When tested in a more objective way, the percentages of rejection show a scalar behavior which nicely fits the gradient judgments.

With this general picture in mind, I suggest that the results of all the experiments reported in this paper should be interpreted as NOT containing clear evidence for syntactic constraints on BrP BSs. Rather, these forms are grammatical (even if sometimes only marginally) and in need of contextual support in order to be used or interpreted. An explicit account for this cannot be given here, but I think that an approach like Pires de Oliveira and Rothstein’s (2011, in press) which does not posit syntactic constraints but makes use of semantic and pragmatic (mis)matches instead is clearly the most promising at the moment. Since in this approach the definite singular and the BS are treated differently and the definite singular is taken to be the canonical form in anaphoric or definite/specific contexts, we get exactly the picture we need: There are two possible forms of which

one is more marked. It will therefore have a more marginal status in the system and only appear occasionally under certain conditions (which remain to be worked out in more detail).

5.3.2 The role and interpretation of grammaticality judgments within an integral approach to BrP BNs

One of the main points of this paper was to show that most of the theoretically interesting uses of BSs explored in the literature are problematic for judgments of well-formedness: there is strong variation and a tendency towards lower judgments and even a non-marginal amount of rejection. This is true for non-prototypical bare mass and count nouns, BSs occurring with episodic or kind predicates as well as in anaphoric/definite/specific contexts. However, as has been shown, this does not warrant premature conclusions about (un)grammaticality. If we accept the view that there are no syntactic restrictions on BSs in BrP, that they denote kinds and licensing takes place via semantic and pragmatic (mis)matches, does that mean that AJTs are not the right method to investigate them because their results are not straightforwardly interpretable? I think that this is not the case. It only means that the next generation of AJTs should be designed with the experiences obtained from the first wave of experiments, the ones discussed in this paper. And there are plenty of cases in need for more objective judgment data.

However, AJTs alone are not sufficient, and neither are reports of introspective judgments of the researcher. Therefore, I argue for a combined approach, making **necessarily** use of (a) corpus studies, (b) acceptability judgment tasks and (c) elicitation/processing tasks. As also shown in chapter 7, studying a phenomenon with different methods is a good way to rule out alternative explanations. In the case of marginal phenomena, the obvious, omnipresent and often quite seductive “alternative explanation” is that such chaotic facts hardly can be part of grammar. This is one reason for the necessity of combining approaches. Another crucial one is that typically, in such cases the results obtained with only one method are difficult to interpret or just deliver no positive evidence at all. In the following three blocks I list the main weaknesses of each method alone which are sufficient to question the results in the case of marginal phenomena. Then I point out how the additional use of other data types compensates these shortcomings in each case. Ideally, if the marginal phenomenon is more than a phantom or an artifact, an interpretable pattern should arise through the combination of the different data types:

- a. Marginal phenomena in corpora: hardly visible, if at all
Possible premature conclusion: ‘The phenomenon does not exist.’
But: Judgment experiments to determine the grammatical status, “controlled recordings” to see how the phenomenon can be boosted in order to obtain more examples.

- b. Marginal phenomena in AJTs: only contradictory judgments
Possible premature conclusion: ‘The phenomenon is ungrammatical.’
But: Examples from corpora can help to design experiments and interpret the results; the same is true for observing the behavior of the phenomenon in controlled recordings²⁰.
- c. Marginal phenomena in elicitation/processing tasks: result of artificial situation
Possible premature conclusion: ‘Speakers are manipulated, no authentic results.’
But: background of judgment experiments and examples from corpora overrule the objection

If a pattern emerges, the method should deliver even more clear-cut results in a second application on the basis of the first results and obviously, a “cyclical” repetition of points (a)-(c) would be the ideal procedure.

Contemporary linguistic theories have reached a level of detail where even certain quite marginal phenomena seem to be highly relevant. If some are, why not all? This would surely be an exaggeration but I think the still widespread basic assumption that marginal phenomena are automatically theoretically less important or even negligible needs to be corrected and that marginal phenomena should only be considered as theoretically insignificant when clear extra-linguistic accounts can be given for them. As long as this is not the case, they should be taken seriously in linguistic theorizing.

5.4 Conclusion

This article is concerned with two main points, which of course are inter-related: One is a better theoretical foundation for the interpretation of AJT results on BrP BSs and the other is the reopening of the discussion of BrP BNs by introducing a phenomenon ignored in most of the literature on this topic, namely the marginal uses of BNs with a definite/specific interpretation.

²⁰ This is not a slight contradiction to (a), as a reviewer suspects. “Hardly visible” is a comparison with contemporary high-level automatic corpus analysis where up to billions of words are scanned and usually many thousands of occurrences are reported in order to provide an objective measure. It also refers to the fact that as yet, there is no method for automatically identifying BrP BSs since this would mean searching for missing articles before nouns in specific semantic contexts. Cf. Wall (in press) for the many challenges a corpus study on BrP BSs has to face. Nonetheless, a collection of some hundreds of (admittedly very different) examples of course allow for first hypotheses that can be tested or for comparison with elicitation patterns from more “artificial” situations.

It has been argued that on the basis of the AJT results reported on BSs, it is not possible to derive conclusions for (un)grammaticality since there is no clear cut-off point. The consequent danger of arbitrary interpretations has been visualized in the cases of BS+kind predicates and specific/definite/anaphoric BSs. As a way out of this dead end, a combination of methods has been suggested, sketched and exemplified by the experiments reported in section 5.2.

From the theoretical perspective, it has been suggested that at the moment Pires de Oliveira & Rothstein's (2011, in press) approach, which posits no syntactic restrictions on BSs and treats all of them as kind-denoting mass terms, is the most promising one since as far as I see, it presents no a priori problem for any kind of data discussed in the literature. Of course many theoretical implications of this approach still need to be developed and empirically verified. Still, the idea of an account exploring semantic and pragmatic (mis)matches rather than syntactic restrictions it is quite appealing as it seems to be really congenial to the vacillating gradient perception of well-formedness observed for syntactically still quite similar patterns.

References

Printed materials used in the experiments:

- Mayer, Mercer. 2003. *Frog, Where Are You?* Reprint. New York: Dial Books.
 Mayer, Mercer. 2003. *A Boy, a Dog, a Frog, and a Friend. Reprint.* New York: Dial Books.
 Mayer, Mercer. 2003. *Frog on His Own.* Reprint. New York: Dial Books.
 Mayer, Mercer. 2003. *Frog Goes to Dinner.* Reprint. New York: Dial Books.

Corpora:

- PEUL. Programa de Estudos sobre o Uso da Língua. Universidade Federal do Rio de Janeiro.
<http://www.lettras.ufrj.br/peul/amostras%201.html>. Accessed 06 Spetember 2013.
 Banco de dados IBORUNA. UNESP São João do Rio Preto.
<http://www.iboruna.iblce.unesp.br/index.php>. Accessed 06 Spetember 2013.

References:

- Amaral, Amadeu. 1920. *O dialeto caipira: gramática - vocabulário*. São Paulo: Huítec.
 Bader, Markus, and Jana Häussler. 2010. Toward a model of grammaticality judgments. *Journal of Linguistics* 46(2): 273-330.
 Bisol, Leda. 2003. External Sandhi in Brazilian Portuguese. *Probus* 15(2): 177-200.
 Braga, João Vinícius de A., Laiza de Sena, Ruan Mariano, and Roberta Pires de Oliveira. 2010. Bare singular and bare mass nouns in Brazilian Portuguese: First results of an empirical survey. *Jornal of Portuguese Linguistics* 9(1): 75-94.
 Bott, Oliver, and Fabian Schlotterbeck. 2009. The processing domain of scope interaction. Poster presented at CUNY 2009, Davis, California.
 Chierchia, Gennaro. 1998. Reference to kinds across languages. *Natural Language Semantics* 6(4): 339-405.
 Dayal, Veneeta. 2011. Bare Noun Phrases. In *Handbook of Semantics* [HSK 33.2], eds. Claudia Maienborn, Klaus von Heusinger, and Paul Portner, 1088-1108. Berlin/Boston: Mouton de Gruyter.
 Dobrovie-Sorin, Carmen, and Roberta Pires de Oliveira. 2008. Reference to Kinds in Brazilian Portuguese: Definite Singulars vs. Bare Singulars. In: *Proceedings of Sinn und Bedeutung 12*, ed. Atle Grønn, 107-121. Oslo: ILOS.

- Featherston, Sam. 2008. Thermometer judgements as linguistic evidence. In *Was ist linguistische Evidenz? Kolloquium des Zentrums Sprachenvielfalt und Mehrsprachigkeit* [ZSMStudien 2], Claudia Riehl, and Astrid Rothe, 69–89. Aachen: Shaker.
- Ionin, Tania, Silvina Montrul, and Hélade Santos. 2011. An experimental investigation of the expression of genericity in English, Spanish and Brazilian Portuguese. *Lingua* 121(5): 963–985.
- Journal of Portuguese Linguistics*. 2010, vol. 9(1). Special Issue: Bare Noun Phrases in Focus (guest editor: Roberta Pires de Oliveira).
- Kabatek, Johannes. 2002. Gibt es einen Grammatikalisierungszyklus des Artikels in der Romania? *Romanistisches Jahrbuch* 53: 56–80.
- Krifka, Manfred, Francis Jeffrey Pelletier, Gregory N. Carlson, Alice ter Meulen, Godehard Link, and Gennaro Chierchia. 1995. Genericity: An introduction. In *The generic book*, eds. Gregory N. Carlson, and Francis Jeffrey Pelletier, 1–124. Chicago: Chicago University Press.
- Müller, Ana. 2000. Sentenças Genericamente Quantificadas e Expressões de Referência a Espécies no Português do Brasil. *Cadernos de Estudos Linguísticos, Campinas* 39: 141–158.
- Müller, Ana. 2002. The Semantics of Generic Quantification in Brazilian Portuguese. *Probus* 14(2): 279–298.
- Müller, Ana, and Fátima de Oliveira. 2004. Bare Nominals and Number in Brazilian and European Portuguese. *Journal of Portuguese Linguistics* 3(1): 9–36.
- Munn, Alan, and Cristina Schmitt. 2002. Bare nominals and the morphosyntax of number. In *Current Issues in Romance Linguistics* [CILT 220], eds. Teresa Satterfield, Christina Tortora and Diana Cresti, 217–231. Amsterdam: John Benjamins.
- Munn, Alan, and Cristina Schmitt. 2005. Number and Indefinites. *Lingua* 115(6): 821–855.
- Pires de Oliveira, Roberta, and Ruan Mariano. 2011. MULHER_F discutiu futebol na festa ontem! Estrutura informational e os nomes nus no PB. *Anais do VII Congresso Internacional da Abralín*, Curitiba: 3744–3756.
- Pires de Oliveira, Roberta, and Susan Rothstein. 2011. Bare singular noun phrases are mass in Brazilian Portuguese. *Lingua* 121(15): 2153–2175.
- Pires de Oliveira, Roberta, and Susan Rothstein. in press. Bare singular arguments in Brazilian Portuguese: perfectivity, telicity, and kinds. In *New perspectives on bare noun phrases in Romance and beyond* [SLCS 141], eds. Johannes Kabatek, and Albert Wall, 189–222. Amsterdam: John Benjamins.
- Pires de Oliveira, Roberta, Josa Coelho da Silva, and Mariana Rublescki Silveira Bressane. 2010. O singular nu denota espécie: uma investigação empírica. *D.E.L.T.A* 26(1): 115–139.
- Schmitt, Cristina, and Alan Munn. 1999. Against the Nominal Mapping Parameter: Bare nouns in Brazilian Portuguese”. In *Proceedings of NELS 29*, eds. Pius Tamanji, Masako Hirotoni, and Nancy Hall, 339–353. Amherst, MA: GLSA.
- Schmitt, Cristina, and Alan Munn. 2002. The syntax and semantics of bare arguments in Brazilian Portuguese. *Linguistic Variation Yearbook* 2: 253–269.
- Wall, Albert. in press. The distribution of definite and specific bare nominals in Brazilian Portuguese. In *New perspectives on bare noun phrases in Romance and beyond* [SLCS 141], eds. Johannes Kabatek, and Albert Wall, 223–253. Amsterdam: John Benjamins.